

**Планирование задач и оптимизация логистических  
операций в облачных производственных системах с  
помощью обучения с подкреплением.**

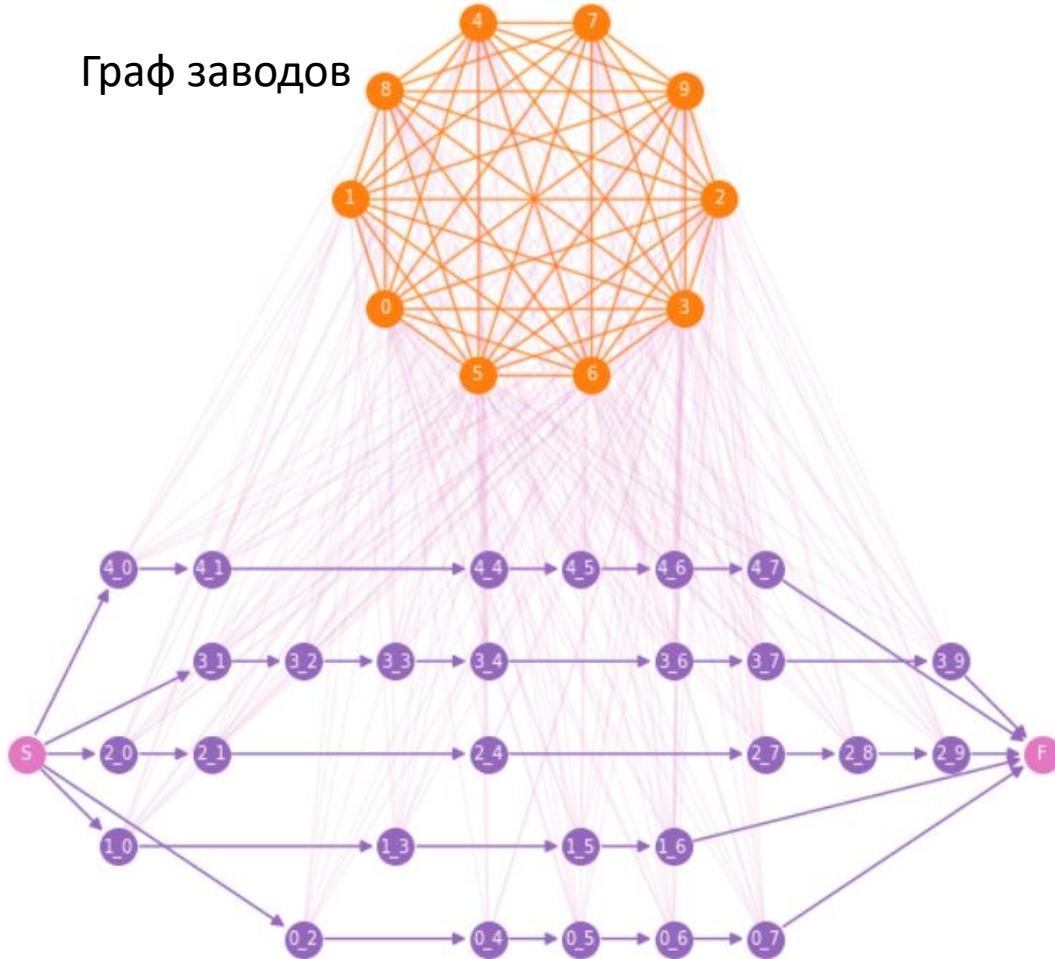
Студент: Журавлёв Дмитрий, Б05-111

Руководитель: Виталий Поздняков,  
младший научный сотрудник АИРИ

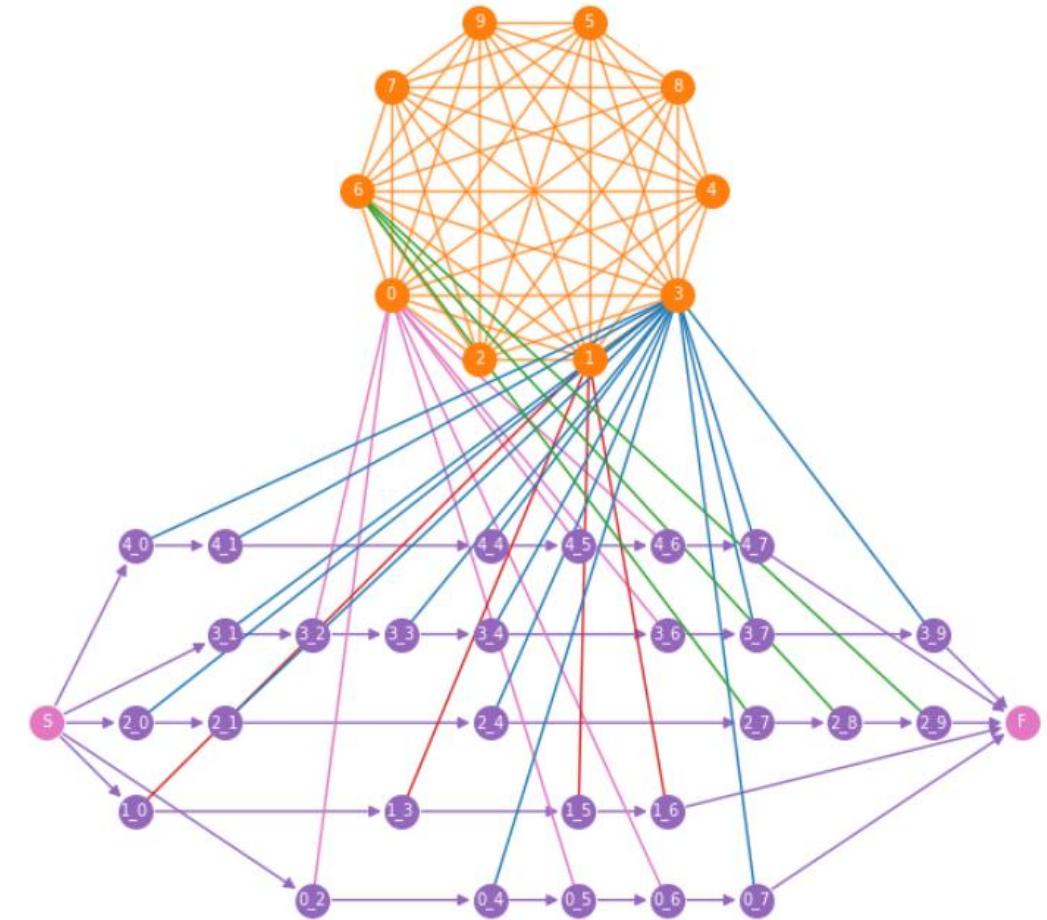
## Постановка задачи

Задача: запланировать производство каждой подзадачи так, чтобы минимизировать расходы на производство и логистику

Граф заводов



Визуализация поставленной задачи



Решенная задача планирования

# Данные

	Operations Mat				
	Operation1	Operation2	Operation3	Operation4	Operation5
Sub-operation1	1	1	0	1	0
Sub-operation2	0	1	0	1	0
Sub-operation3	1	1	1	0	1
Sub-operation4	1	0	1	1	0
Sub-operation5	1	1	0	1	1
Sub-operation6	1	1	0	1	1
Sub-operation7	1	1	1	1	0
Sub-operation8	1	1	1	0	1
Sub-operation9	1	1	1	1	1
Sub-operation10	1	1	0	0	1

	Cut Distances Mat					
	city1	city2	city3	city4	city5	...
city1	0	1633,84293	611,70304	89,7207668	1130,56529	...
city2	1633,84293	0	1739,44141	1718,00095	2678,55942	...
city3	611,70304	1739,44141	0	649,74271	1528,56031	...
city4	89,7207668	1718,00095	649,74271	0	1041,86721	...
city5	1130,56529	2678,55942	1528,56031	1041,86721	0	...
...	...	...	...	...	...	...

	Cut Times Mat					
	city1	city2	city3	city4	city5	...
Sub-operation1	Inf	3,78806541	6,2787035	6,53023044	5,1937218	...
Sub-operation2	Inf	Inf	3,17855839	3,15916423	4,90779229	...
Sub-operation3	3,63493408	Inf	Inf	4,38461492	6,82758394	...
Sub-operation4	Inf	5,42687824	Inf	3,23085695	6,97599951	...
Sub-operation5	6,16179623	Inf	6,39367577	3,48565891	3,93436302	...
Sub-operation6	3,48770202	3,70943169	6,78870065	Inf	5,44882198	...
Sub-operation7	4,39249109	5,10880641	6,71566234	6,47414311	5,227931	...
...	...	...	...	...	...	...

	Cut Costs Mat					
	city1	city2	city3	city4	city5	...
Sub-operation1	Inf	31,0329157	65,9018489	69,4232262	50,7121052	...
Sub-operation2	Inf	Inf	22,4998175	22,2282992	46,709092	...
Sub-operation3	28,8890771	Inf	Inf	39,3846089	73,5861752	...
Sub-operation4	Inf	53,9762954	Inf	23,2319973	75,6639931	...
Sub-operation5	64,2651472	Inf	67,5114608	26,7992247	33,0810823	...
Sub-operation6	26,8278283	29,9320437	73,0418091	Inf	54,2835077	...
Sub-operation7	39,4948753	49,5232898	72,0192728	68,6380036	51,191034	...
...	...	...	...	...	...	...

[Датасет](#): Dataset for logistics and manufacturing service composition, 2021.

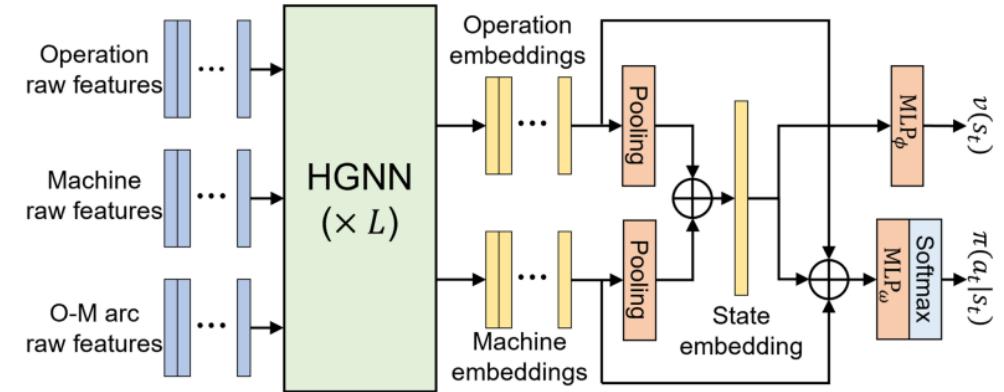
## Статья авторов

Jalar Delaram, Omid Fatahi Valilai

# Обзор текущих решений

1. Статья. Гибкое планирование работы с помощью графовой нейронной сети и обучения с глубоким подкреплением

(Вэнь Сон; Синъян Чен; Цицян Ли; Чжигуан Цао)

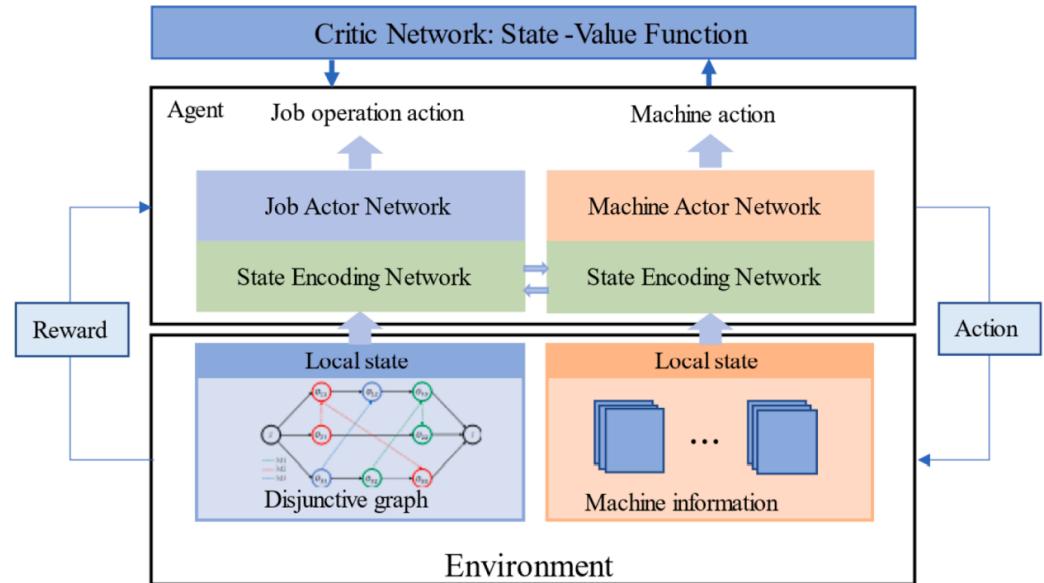


Network architecture.

2. Статья. Многодействующая система обучения с глубоким подкреплением для решения задачи гибкого планирования работы магазина

(Кун Лай, Пинг Го, Вэнь чао Чжао и др.)

Отметим, что оба решения не учитывают затрат на логистику и имеют сложную архитектуру.



Multiple actor-critic architecture for a multi-action space scheduling problem.

# Мотивация моего исследования

- 1) GNN без RL, но с затратами на логистику -> добавить RL  
Энкодер

Свертка в вершины S

- $h_i^s = W^s S_i$
- $h_{ij}^{ss} = W^{ss} [S_i \| SS_{ij}]$
- $h_{ij}^{os} = W^{os} [O_i \| OS_{ij}]$
- $e_{ij}^{os} = \text{LeakyReLU} \left( a^\top \begin{bmatrix} h_{ij}^{os} \\ h_j^s \end{bmatrix} \right)$
- $e_{ij}^{ss} = \text{LeakyReLU} \left( a^\top \begin{bmatrix} h_{ij}^{ss} \\ h_j^s \end{bmatrix} \right)$
- $\alpha_i^{os} = \text{Softmax}_{(i,j) \in E_{os}} e_{ij}^{os}$
- $\alpha_i^{ss} = \text{Softmax}_{(i,j) \in E_{ss}} e_{ij}^{ss}$
- $z_j = \sum_{(i,j) \in E_{ss}} \alpha_i^{ss} h_{ij}^{ss} + \sum_{(i,j) \in E_{os}} \alpha_i^{os} h_{ij}^{os}$

свертка в вершины O

- $x_j = W^o \text{ReLU} \left[ \sum_{(i,j) \in E_{oo}} W^{\text{in}} O_i \| W^{\text{self}} O_j \| \sum_{(j,i) \in E_{oo}} W^{\text{out}} O_i \right]$

## Декодер

Софтмакс на ребрах

$$p_{ij} = \sigma(z_i^\top x_j)$$

- 2) Есть жадный алгоритм ->  
RL даже без GNN должен работать лучше

Operation, sub- Operation, City	Optimal-S	Estimated- S	Optimi- zation- T	Estima- tion- T
5, 10, 10	7652,32	12521,69	88,5	0,0007
10, 10, 10	14200,75	23082,48	315,21	0,0019

*Сравнение точного и жадного решения.*

## Проделанные шаги исследования:

Изучение архитектуры GNN. Решение учебного задания.



Изучение архитектуры actor-critic. Решение учебного задания.



Запуск кода опорных статей. Существенное изменение траектории исследования.



Разработка Q-learning решения. Численные эксперименты.

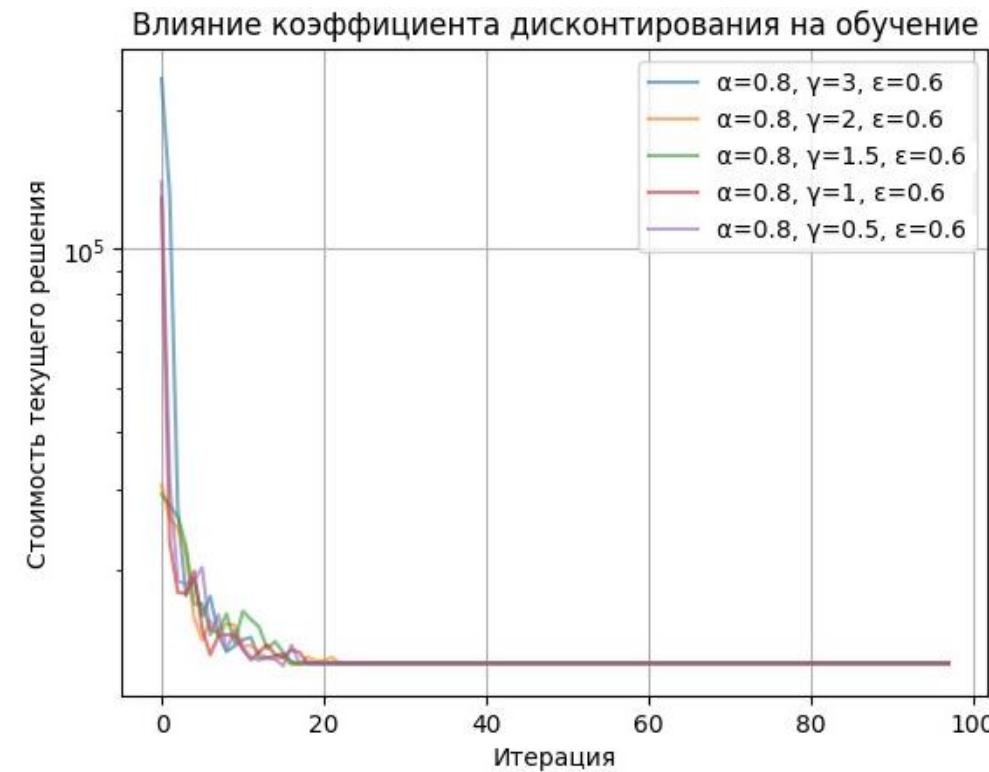
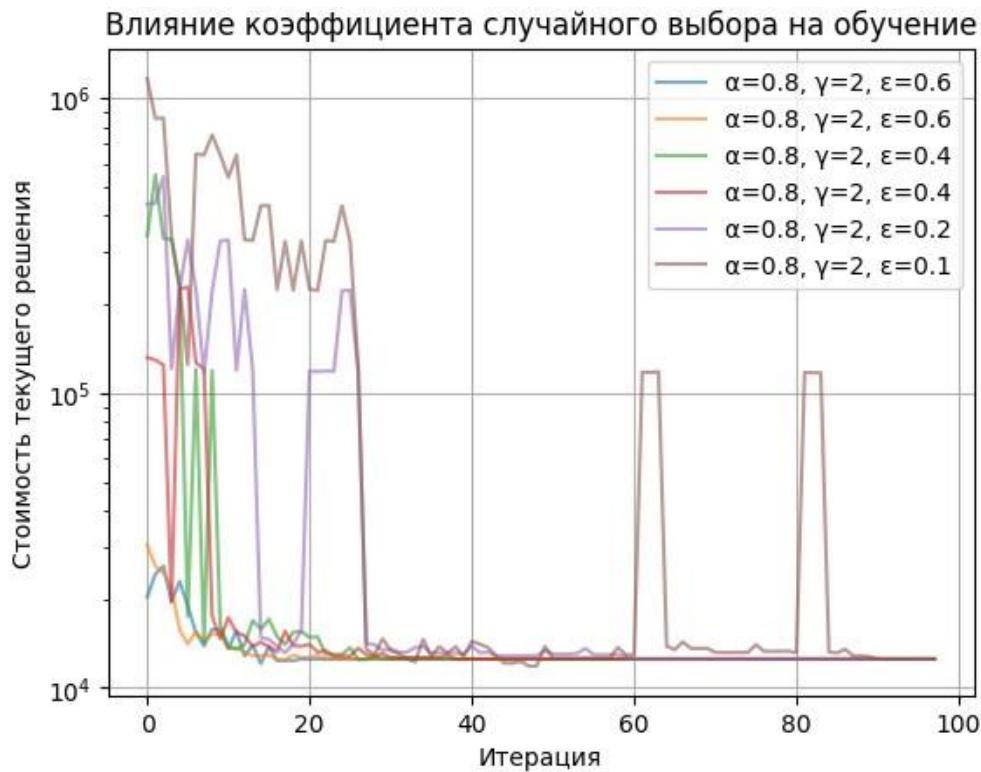


Обоснование результата. Исследование пространства решений.

Git: [https://github.com/Zhyravlev-Dmitry/Q-learning\\_for\\_logistics.git](https://github.com/Zhyravlev-Dmitry/Q-learning_for_logistics.git)

# Q-learning

График обучения при различных гиперпараметрах



Уравнение пересчёта:  $Q^{new}(S_t, A_t) = (1 - \alpha) \cdot Q(S_t, A_t) + \alpha \cdot (R_{t+1} + \gamma \cdot \min_a Q(S_t, a))$ , где

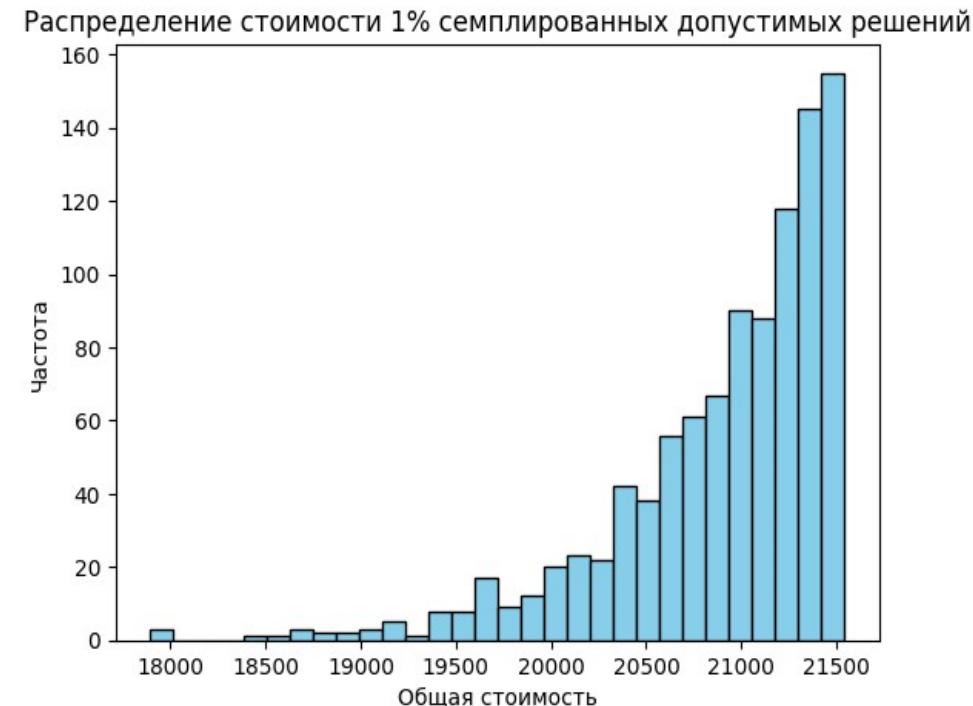
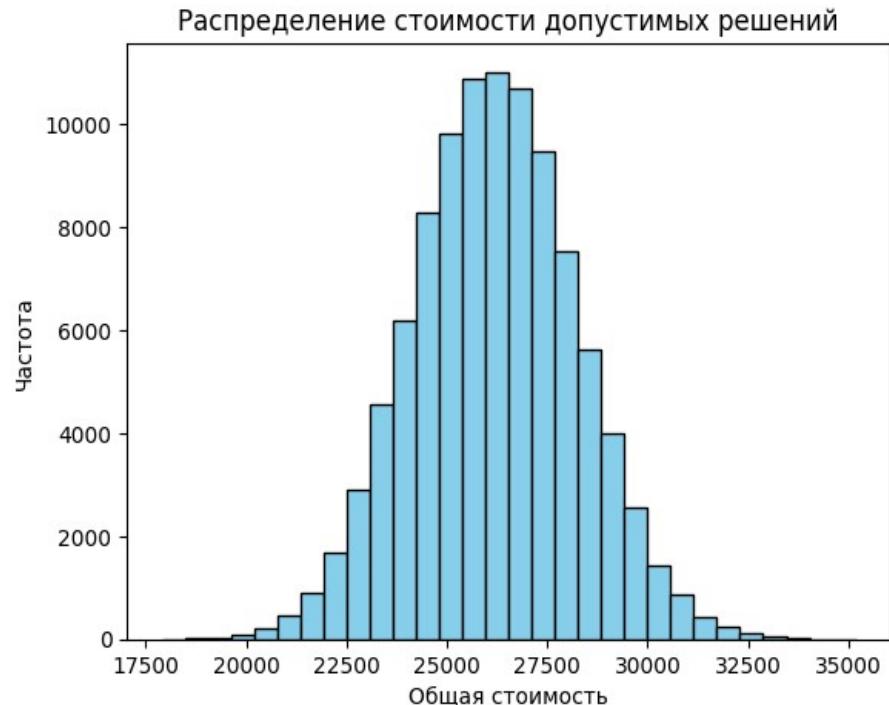
$Q$  – функция ожидаемых вознаграждений,  $\alpha$  – скорость обучения ( $0 < \alpha \leq 1$ ),  $R_{t+1}$  – стоимость производства при переходе из состояния  $S_t$  в  $S_{t+1}$ ,  $\gamma$  – коэффициент дисконтирования,  $S_t$  – состояние, определяемое текущей операцией, субоперацией и городом,  $A_t$  – возможные города производства следующей за текущей субоперации.

# Сравнительная таблица и обоснование результата

Операции, субоперации, города	Точное решение	Жадный алгоритм	Q-learning	GNN
5, 5, 10	7652	12522	12522	12369
10, 10, 10	14200	23082	23082	22574
5, 10, 20	5866	9414	9414	9625

*Сравнение результата работы разных алгоритмов*

*График стоимости случайных решений и субоптимальных случайных решений.  
Количество семплов: 100 000*



## Выводы:

- В работе разработано решение задачи на основе Q-learning
- Решение на основе Q-learning не дает значительного улучшения стоимости вследствие специфики пространства возможных решений и принципа работы алгоритма
- Решение на основе Q-learning в данной задаче почти всегда сходится к жадному решению и незначительно зависит от гиперпараметров.
- Результат, полученный на основе Q-learning сравним с результатом, полученным на GNN без RL подхода на малом количестве эпох